DOCUMENT RESUME

ED 276 725 SP 028 405

AUTHOR Melnick, Susan L.; Raudenbush, Stephen W.

TITLE Influence of Pupils' Gender, Race, Ability, and Behavior on Prospective and Experienced Teachers' Judgments about Appropriate Feedback. Research Series No. 175.

INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Oct 86
CONTRACT 400-81-0014
NOTE 41p.
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($3.50).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Elementary Education; Feedback; Racial Bias; Sex Bias; Student Attitudes; Student Behavior; *Student Characteristics; *Teacher Attitudes; Teacher Behavior; *Teacher Response; *Teacher Student Relationship

ABSTRACT

Because certain teacher behaviors which maximize student engagement in learning have been found to increase student achievement, the sources of those behaviors provide an important focus for research on teaching. The finding that teachers treat children of different gender, race/ethnicity, or ability in ways that may have deleterious consequences for subgroups of children has been a recurrent theme in classroom research. One important source of teacher behavior is teacher judgments about how best to respond to students during public recitation. This paper examines how student gender, race/ethnicity, behavior, and ability influence prospective and practicing teachers' judgments about how to respond to students who answer questions incorrectly during hypothetical episodes of reading recitation. During the sessions, student behavior and ability significantly influenced teachers' judgments: Teachers were more "sustaining" toward on-task students than off-task students and more sustaining toward low-ability students than high-ability students. There were no main effects of gender or race, but interactions including behavior, ability, gender, and race suggested that student characteristics influence the purposes for which sustaining behaviors are used and the messages they convey. (A three-page reference list concludes the document.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Research Series No. 175

INFLUENCE OF PUPILS' GENDER, RACE, ABILITY, AND BEHAVIOR ON PROSPECTIVE AND EXPERIENCED TEACHERS' JUDGMENTS ABOUT APPROPRIATE FEEDBACK

Susan L. Melnick
Stephen W. Raudenbush

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. L. Melnick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Research Series No. 175

INFLUENCE OF PUPILS' GENDER, RACE, ABILITY, AND BEHAVIOR ON PROSPECTIVE AND EXPERIENCED TEACHERS' JUDGMENTS ABOUT APPROPRIATE FEEDBACK

Susan L. Melnick
Stephen W. Raudenbush

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824-1034

October 1986

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Office of Educational Research and Improvement, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office or the Department. (Contract No. 400-81-0014)

Institute for Research on Teaching

The **Institute for Research on Teaching** was founded at Michigan State University (MSU) in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second five-year contract to MSU. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Co-Directors: Jere E. Brophy and Andrew C. Porter

Associate Directors: Judith E. Lanier and Richard S. Prawat

Editorial Staff
Editor: Sandra Gross
Assistant Editor: Sally B. Pratt

Abstract

Because certain teacher behaviors which maximize student engagement in learning have been found to increase student achievement, the sources of those behaviors provide an important focus for research on teaching. The finding that teachers treat children of different gender, race/ethnicity, or ability in ways that may have deleterious consequences for subgroups of children has been a recurrent theme in classroom research. One important source of teacher behavior is teacher judgments about how best to respond to students during public recitation. This paper examines how student gender, race/ethnicity, behavior, and ability influence prospective and practicing teachers' judgments about how to respond to students who answer questions incorrectly during hypothetical episodes of reading recitation. During the sessions, student behavior and ability significantly influenced teachers' judgments: Teachers were more "sustaining" toward on-task students than off-task students and more sustaining toward low-ability students than high-ability students. There were no main effects of gender or race, but interactions including behavior, ability, gender, and race suggested that student characteristics influence the purposes for which sustaining behaviors are used and the messages they convey.

INFLUENCE OF PUPILS' GENDER, RACE, ABILITY, AND BEHAVIOR ON PROSPECTIVE AND EXPERIENCED TEACHERS' JUDGMENTS ABOUT APPROPRIATE FEEDBACK[1]

Susan L. Melnick and Stephen W. Raudenbush[2]

For nearly two decades research on teacher effects has demonstrated that certain teacher behaviors which maximize student engagement have been found to increase student achievement (Brophy & Good, 1986). This research is reinforced by the findings of the research on teacher expectations and their mediating mechanisms (Cooper & Good, 1983; Dusek & Joseph, 1985; Rosenthal & Rubin, 1978; Smith, 1980). In general, the behaviors which reflect positive performance expectations are similar to those which have been found to maximize student engagement and achievement.

Our review of observational research indicates that teachers tend to differentiate their behavior during classroom interactions in ways that are likely to have negative effects on some students' opportunities for learning and hence on their achievement. Specifically, teachers have been found to differentiate their behaviors in ways that create more opportunities to learn for boys than for girls, for majority than for minority youngsters, and for those presumed to be of high ability than for those viewed as "lows."

---

[1]Portions of this work were presented at the annual meeting of the American Educational Research Association, New Orleans, April 1984.

[2]Susan L. Melnick is coordinator of the Interactive Decision Making and the Prospective Teacher Project and an assistant professor of teacher education at Michigan State University. Stephen W. Radenbush is an assistant professor in MSU's Department of Counseling, Educational Psychology and Special Education. The authors acknowledge Donna Wanous Bullock's contribution to the initial design and conduct of this study. Thanks also to Randall P. Fotiu for his assistance in data analysis and to Jere Brophy and Andrew Porter for their comments on an earlier draft of this manuscript.

Although no definitive explanation for the differential treatment has been observed, a number of possible interpretations are possible. Some might attribute such differentiation to teacher bias (Dusek & Joseph, 1985). Others have argued that teachers are accurate judges of student ability, achievement, and behavior and that those judgments are the primary source of teacher expectations and actions (Brophy, 1983). Thus, one possible explantion is that teacher behaviors observed during classroom interactions simply reflect teachers' best judgments about what is pedagogically appropriate for the individual student with whom they interact. It may be, however, that under the press of classroom life, teachers' decisions to sustain or terminate interaction with a specific student are molded more by the demands of managing the group or by the need to maintain the flow of the lesson than by judgments about what is most appropriate pedagogically for a particular youngster (Cooper, 1977; Doyle & Carter, 1984).

One source of difficulty in assessing the alternative explanations is that little is known about how students' behavior and characteristics influence teachers' judgments about what is pedagogically most appropriate during interactive teaching episodes. Thus, it is not clear if teachers think, for example, that they should sustain "highs" more than "lows" or boys more than girls when the student in question responds incorrectly. As a result, it is difficult to determine if such behaviors observed in the context of the classroom reflect teachers' judgments about what is most appropriate for a given student or if such behaviors result from a myriad of other factors--unconscious bias, the press of classroom events, group needs, or the need to push ahead with the curriculum.

This study focuses on the effect of student gender, race/ethnicity, ability, and behavior on the pedagogical judgments teachers make when responding to individual students during critical moments in recitation. First, we compared the effects of student characteristics on teachers' judgments to past findings. Second, we examined the extent to which teachers differentiate their judgments on the basis of student gender, race/ethnicity, behavior, and ability. Last, we explored possible consequences of such differentiation for practice and future research.

Effects of Student Characteristics on Teacher Behavior and Expectations

Gender

According to Bossert (1981), some evidence supports the hypothesis that girls and boys receive differential teacher treatment. Teachers have been found to provide boys with more criticism, academic praise, help, and overall attention and to refer boys more often than girls for disciplinary action (Blumenfeld, Hamilton, Wessels, & Falkner, 1977; Brophy & Good, 1974; Etaugh & Harlow, 1975). Moreover, in mathematics teachers have been found to provide girls with fewer academic contacts, fewer complex or abstract questions, and less instruction on how to do problems for themselves (Leinhardt, Seewald & Engel, 1979). Other studies, however, (e.g., Brophy & Everston, 1981) suggest patterns of more favorable teacher perceptions of girls than boys, despite the fact that boys appear to receive more academic and personal contact, positive and negative.

Although it appears, overall, that boys have greater salience in classrooms and that girls are more likely to be invisible, Bossert (1981) has criticized the reliance of most studies on global ratings of frequency of occurrence,

without attention to variations across subject matter and instructional activity or to student mediating effects. Further, Good and Findley (1985) found that, of five studies of differential teacher expectations based on student gender, predictions were largely dependent upon the specific situation. In a meta-analysis of 20 studies, Dusek and Joseph (1985) found that teachers' expectations were somewhat higher for girls than for boys with respect to measures of social/personality development. (No gender effect with respect to expectations for academic achievement was reported.) In summary, substantial evidence suggests gender-based differential teacher treatment exists in classroom interaction and expectations. Research has not examined adequately the influence of gender on teacher judgments about appropriate responses.

Race/Ethnicity

Irvine (1985) cites a number of experimental studies that indicate that white students receive more favorable communication than blacks. In addition, studies of teacher ratings of their perceptions of black and white students (e.g., Aloia, Maxwell, & Aloia, 1981) generally conclude, on average, that teachers have more negative attitudes and beliefs about black children regarding potential for success, ability, and behavior. Naturalistic studies yield similar findings: Black students receive more negative academic and behavioral feedback than white students (Aaron & Powell, 1982; Brophy, 1983; Brophy & Good, 1974; Simpson & Erickson, 1983). In a review of 16 studies comparing white and black students, Baron, Tom, and Cooper (1985) found that teacher expectations favored white students. They also found that teachers held higher expectations for white than for Mexican-American students, whereas expectations for blacks and Mexican-Americans did not differ. The bases for

such expectations for whites, Dusek and Joseph (1985) argue, "likely reflect stereotypic (prehaps prejudicial) expectancies for social behaviors. In the absence of more academically relevant information, teachers may rely on this type of knowledge, imperfect though it may be, about students when forming initial impressions and expectations" (p. 243).

Student Behavior

Some evidence supports the contention that teachers hold higher expectations generally for students who behave appropriately in the classroom (e.g., Purgess, 1979; Willis & Brophy, 1974). Adams and LaVoie (1974) have suggested that the child whom the teacher perceives as not conforming to behavioral expectations is viewed as having less potential for success and lower ability. In addition, some studies indicate a possible interaction of teachers' perceptions of negative behavior with student race or ethnicity. Several studies reviewed by Irvine (1985) suggest that teachers regard the behavior of black children as more deviant than white children. Roberts, Hutton, and Plata (1985) found that teachers perceived the behavior of minority students as problematic. Black students were seen as demonstrating more behavior problems than whites, but teachers rated the behavior of Hispanic students less favorably than that of either black or white students. Their findings were consistent with those of Argulewicz, Elliott, and Hall (1982) who found that Hispanic students demonstrated fewer learning and motivational behaviors and were thus perceived to avoid more frequently important teacher and peer interactions. Whereas teachers disagree about what specifically constitutes inappropriate behavior, it appears that student race/ethnicity tends to mediate the perceived appropriateness of behavior in the classroom.

Student Ability

In extensive reviews of studies of teacher-student interaction, Brophy and Good (1974) and Brophy (1983) found substantial evidence of differential treatment based on student ability. In general, teachers have been found to wait less time for low-ability students than for high-ability students to answer questions and to give low-ability students the answer more often than high-ability students. When a low-ability student responds incorrectly, the teacher has the option of asking another student, thus terminating the interaction, or of staying with the original student. Teachers have been found to exercise the "ask another student" option with higher probability when the student who answers incorrectly is perceived to be of low-ability.

Teachers have also been found to give low-ability students inappropriate reinforcement or feedback for public responses. There is also a higher probability that teachers will criticize lows more often for failure and praise them less frequently for success. Teachers have been found to pay less attention to lows, to interact with or call on them less, and to fail to give them the benefit of the doubt in borderline cases. In addition, teachers have been found to give low-ability students less opportunity to practice lessons independently, give more intrusive instruction, and provide briefer and less informative feedback on their questions.

In a series of studies, Brophy and Good (1974) linked teachers' naturalistic expectations for different students to differential patterns of teacher-student interaction. For example, consistent with earlier findings for low-ability students, those who were perceived by teachers to be of high ability were more likely to receive praise when they succeeded and less likely to receive criticism when they failed. High-ability students received more

specific feedback regarding the correctness of their responses than low-ability students and were more likely to have interaction sustained when they hesitated during reading or when they failed to answer a question or answered incorrectly.

The Present Study

This study was designed to investigate the effects of student gender, race/ethnicity, behavior, and ability on teacher judgments about the appropriateness of responses to students. Although most previous research on student characteristics has involved either predictions about future outcomes or documentation of classroom interaction, this study used a policy-capturing methodology to determine the effects of student characteristics and behavior on teacher judgments under hypothetical circumstances. Although this method is not designed to uncover the mental operations leading to judgments (Floden, Porter, Schmidt, Freeman, & Schwille, 1981), the policy-capturing approach permits the isolation and examination of certain variables related to judgment outcomes confounded by the daily press of actual classroom teaching practice. By studying simulated rather than actual classroom situations, the researchers were able to control student characteristics and behaviors experimentally to facilitate causal inferences about their relationship to teacher judgments.

For two decades research on teaching has clearly pointed to relationships between student engagement and achievement. In addition, this research has documented both the importance of appropriate corrective feedback for learning (e.g., Cooper, 1977; Fisher, Berliner, Filby, Marliave, Cahen, & Dishaw, 1980; Rosenshine, 1976; Zahorik, 1968) and the salience of public recitations as a forum for communication of expectations and display of student knowledge (e.g., Brophy & Good, 1974; Good, 1970; Mendoza, Good, & Brophy, 1972; Rowe, 1974).

Given these findings, the ubiquitous question-answer-feedback pattern provides an ideal situation to determine how selected student characteristics and behavior might influence teacher decisions broadly characterized as sustaining or terminating interaction. Thus, this study focuses not on how teachers behave during actual teacher-student interactions but on their judgments about how to behave under hypothetical interactive teaching moments.

Variables and Hypotheses

This study was one of a projected series designed to consider the kinds of information about students that teachers attend to when making interactive teaching decisions. Its purpose was to determine how selected student assessment and demographic data influence teachers' perceptions about how particular students should be treated during a small-group reading activity. Assessment data consisted of two types of information: (a) a student's ability in reading (above average, average, below average), as represented by her/his score on a standardized reading test and the reading grade s/he received on the last report card, and (b) the student's classroom behavior (on or off task) as observed by the teacher during the particular lesson. The student demographic characteristics refer to the race/ethnicity (black, white, Hispanic) and gender of the student.

The specific interactive decisions refer to the teacher's instruction when a student gives the wrong answer to a question during a reading lesson. The study was designed to explore if subjects' interactive decisions were influenced by the type of assessment data provided or by the descriptions of students' demographic characteristics, if there was an interaction effect between the assessment data and demographic characteristics, and if the decisions of prospective and experienced teachers differed.

The primary focus of the study was the influence of student characteristics on teachers' judgments about whether to sustain or terminate interaction. Given the findings from previous research, it was anticipated that teachers would clearly differentiate their behavior toward on- versus off-task students. Pedagogical principles suggest, for instance, that a student who is on task but responds incorrectly should be encouraged through some sustaining behavior (e.g., providing a clue or rephrasing) or by providing appropriate corrective feedback. In contrast, off-task behavior should be extinguished, or, perhaps more positively, transformed into on-task behavior.

Examples of inappropriate teacher responses might include praising off-task behavior and ignoring or criticizing on-task behavior. Terminating interactions with on-task students, though perhaps not entirely inappropriate, has been found to convey the teacher's low expectancy for the student's success, especially when students of low status or achievement are routinely singled out for such terminating interaction (e.g., Brophy, 1983). At the same time, when teachers differentially sustain such low-status students more than higher status students, the message conveyed might be that such students need help. Thus the meaning of teacher responses may depend on whether the responses are consistent or differential.

More interesting than the main effect of task behavior (on-task vs. off-task) was the possible main effect of gender, race/ethnicity, and ability, and especially the interaction effects of task behavior with those factors included. For instance, if teachers were found to differentiate their responses on the basis of gender or race/ethnicity, the data would provide support for the notion that the teachers had formed differential expectations on the basis of gender or race. Similarly, statistical interaction effects between these demographic characteristics and task behavior would indicate that

teachers believe that their responses to on- or off-task students ought to depend on the student's gender or race. For instance, suppose teachers provided more negative feedback to off-task whites than off-task blacks. This behavior could be interpreted as evidence that the teachers expected off-task behavior of blacks but not of whites and could thus be construed as evidence of racial stereotyping. Similarly, a tendency to sustain on-task girls more than boys could alternatively be construed as evidence that more could be expected of girls or that girls need more help than boys.

Notice that main effects of demographic variables or interactions between those variables and task behavior would support a different interpretation for similar effects observed under actual teaching conditions. In actual classroom interaction, such behavior might betray unconscious or unintended bias. In this simulation study, such effects would tend to reveal that stereotyping occurs on a more conscious level, that is, that teachers believe they ought to behave in ways that discriminate on the basis of such demographic variables. Main effects of ability or interactions between ability and task behavior reveal something about teachers' notions of ability. For instance, encouragement of off-task behavior for low-ability students would reflect not only a low estimate of the student's present status but also, perhaps, a negative prediction about future intellectual growth.

This study also compared the decisions of prospective and experienced teachers. Significant differences between the two groups might suggest that the "real world" of teaching has taught experienced teachers to view teaching differently than novice teachers or that a kind of "cohort effect" was operating. Such an effect might be a result of age differences or differences in the teacher education curriculum over time. No difference between the

prospective and experienced teachers would suggest either similar training, or more likely,that the two groups share an overarching, common social ideology or set of beliefs which prescribe or encourage a common set of interactive judgments.

Method

Subjects

Subjects for the study consisted of 108 prospective and experienced teachers affiliated with two universities, one in the Midwest and one in the Southeast. The prospective teacher sample consisted of 54 volunteer subjects (51 female, 3 male) enrolled in preservice teacher education programs at the two universities. The experienced teacher sample consisted of 54 volunteers (45 female, 9 male) drawn from the pool of master's degree candidates and cooperating teachers from the two sites.

Experimental Procedure

In an attempt to make the hypothetical situations as realistic as possible, 5 experienced elementary teachers helped generate examples of on-task versus off-task behaviors and a list of possible interactive decisions given the particular instructional situation described. Descriptions of 36 hypothetical third-grade students were then constructed by systematically varying four information cues: the student's reading ability level, task behavior, race/ethnicity, and gender. In an attempt to minimize socially desirable responses to the racial variable, these descriptions were further divided into three sets of 12 vignettes, each set representing black, white, or Hispanic students. Thus the between-subjects design involved race/ethnicity of depicted student (black, white, Hispanic) by experience of teacher (experienced vs.

prospective). The within-subjects design included three crossed factors: ability (above average, average, below average), gender, and task behavior (off, on) of the depicted student.

Each vignette consisted of the following: (a) a description of a student's reading ability level, task behavior, race, and gender; (b) a description of a specific instructional interaction in a small-group reading lesson; (c) an indication that the teacher was asking a question; and (d) an indication that the student's response to the question was incorrect. The following vignette is illustrative:

> Jody is a white female of above average ability. She took a standardized achievement test two months ago and scored several months above grade level in reading. Yesterday she took home her report card indicating "Excellent" in reading. At the moment, you are discussing a story the group has just finished reading. Jody has been following along, and she volunteers to answer the next question. You call on her and she gives the wrong answer.

Subjects participated in the study in groups either during class time or after school. At the beginning of each session, subjects were given a brief introduction to the study, indicating only that its purpose was generally to investigate the types of decisions made by prospective and experienced teachers during small-group reading instruction. Each subject then received a detailed description of the hypothetical classroom and a set of vignettes representing only one racial/ethnic group. Within that group, subjects received the 12 vignettes (3 levels of ability x 2 levels of behavior x 2 levels of gender) arranged in random order, and the packets were distributed randomly to the subjects in the group setting. Each subject was directed to read the description of the classroon in his or her packet and then make a decision about each student depicted in the vignettes as if he or she were the teacher in this particular classroom. Responses were collected in one session of one hour or less, with an average completion time of 40 minutes.

Dependent Variables

Subjects were directed to read each vignette independently and to make 12 decisions for each vignette before going on to the next one and without looking back to previous decisions. Subjects were asked to rate separately 12 possible instructional decisions on a 5-point Likert-type scale ranging from "virtually certain to do" (coded 1) to "virtually certain not to do" (coded 5). Teachers were given 12 possible responses to complete the sentence beginning "I would." Possible responses are given in the following chart.

	Sustaining	Terminating	Possibly Sustaining or Terminating
1. Provide specific negative feedback to the student		X	X
2. Repeat the question	X		
3. Ask another student to answer the question		X	
4. Provide the student with a specific clue	X		
5. Rephrase the question	X		
6. Provide specific positive feedback to the student	X		
7. Summarize the material immediately preceding the question	X		X
8. Answer the question myself		X	
9. Ask another student to volunteer to help		X	
10. Ask the same student a new question	X		
11. Capitalize on the student's response by digressing	X		X
12. Provide no feedback		X	X

These decisions may be broadly characterized as sustaining (2, 4, 5, 6, 7, 10, 11) or terminating (1, 3, 8, 9, 12) responses. The teacher could also respond in ways which could occur in the context of either sustaining or terminating the interaction (1, 7, 11, 12). The responses were not viewed as mutually exclusive. Rather, the subjects were asked to rate the likelihood of each of the 12 possible decision options for each vignette. The 12 decisions were presented to the subjects in random order.

Analyses

For each of the 12 decision options the design constituted a split-plot factorial design (Kirk, 1982) with two between-subjects and three within-subjects factors. Two alternative analyses were performed: a univariate mixed-model analysis of variance (ANOVA) approach (Kirk, 1982) and a multivariate repeated measures approach, wherein the responses to each of the 12 vignettes constitute a 12-dimensional vector and the analysis involves a design over the measures (Bock, 1975). Results of these two analyses were essentially identical; the more familiar univariate ANOVA results are reported here. The complexity of the five-factor design (i. e., a 2 x 3 x 2 x 3 x 2 design with repeated measures on the last three factors) allowed up to 71 orthogonal contrasts for each dependent measure. Two methods of testing specific comparisons were employed to render the results theoretically interpretable and to protect against Type I errors (when H_0 or the tested hypothesis is falsely rejected).

First, a limited number of a priori contrasts were computed and the significance levels adjusted by means of the Bonferroni method (Seber, 1977). Specifically, main effects contrasts were computed as follows for each of the 12 dependent measures. For each two-level factor (task behavior, gender,

and teacher experience), one contrast was computed. For the effect of ability, two orthogonal contrasts were computed: a "linear" contrast, which compares high- and low-ability means, and a "quadratic" contrast, which compares the middle-ability mean to the combined high- and low-ability means. For race, two orthogonal contrasts were also computed: a majority (i.e., white) vs. minority (combined black and Hispanic) comparision, and a black vs. Hispanic comparison. Orthogonal a priori contrasts for the two-way interactions were defined by the relevant product terms of the main effects contrasts.

Second, higher order contrasts were investigated by means of the Scheffe method, which offers strong protection against Type I errors. For the purposes of adjusting significance levels by means of the Bonferroni and Scheffe approaches, the sources of variation contributing to each omnibus F-test were considered a "family," and the number of means constituting such a family was the number used to adjust the critical values of F (see Kirk, 1982) at the present significance level.

In addition to tests of significance, the magnitude of the contrasts of significant differences was calculated in standard deviation units. The effect size estimate, or d-index (Cohen, 1969), is defined as the difference between the means of two experimental conditions divided by the standard deviation of the variable. Thus the statement that $d = 1.00$ indicates that an observed mean difference between treatment conditions was equivalent to one standard deviation on the decision option variable and constitutes an unusually large effect. Effect sizes of .20 and .50 would indicate small and moderate effect sizes respectively (Cohen, 1969). For a complex contrast,

d has the same interpretation if the estimated contrast is divided by the standard deviation and if the sum of the absolute values of the contrast coefficients is two.

Results

Main Effects

Task behavior. As indicated above, students who responded incorrectly to academic questions in each vignette were described as behaving in ways which could be characterized as either "on-" or "off task." This "task behavior" strongly influenced the subjects' judgments about how to respond (Table 1). Subjects were substantially more likely to "sustain" on-task students than off-task students, either by providing a clue, $d = .99$, $F\ (1,204) = 208.07$, $p < .001$ or rephrasing the question, $d = .88$, $F\ (1,204) = 139.15$, $p < .001$. They were more likely to "terminate" off-task students than on-task by asking another student to answer the question, $d = -.60$, $F\ (1,204) = 84.57$, $p < .001$; more likely to provide negative feedback, $d = -1.11$, $F\ (1,204) = 203.72$, $p < .001$; and less likely to provide postive feedback, $d = .98$, $F\ (1,204) = 220.09$, $p < .001$, for off- than for on-task students. Smaller statistically significant effects included subjects' inclination to provide no feedback more often for off-task than for on-task students, $d = -.19$, $F\ (1,204) = 9.19$, $p < .01$, and to summarize for on-task students more than for off-task students, $d = .17$, $F\ (1,204) = 91.33$, $p < .01$.

Teachers' judgments about the appropriateness of the other five responses--repeating the question, asking a new question, providing the answer themselves, asking another student to volunteer to help, or digressing--were

Table 1

Effects of Task Behavior

	On-task mean	Off-task mean	Effect size
Sustaining behaviors			
Provide a clue	1.79	2.91	.99**
Rephrase the question	1.60	2.58	.88**
Summarize	2.86	3.07	.17*
Terminating behavior			
Ask other student	3.62	2.92	-.60**
Feedback			
Positive feedback	2.11	3.43	.98**
Negative feedback	4.59	2.16	-1.11**
No feedback	4.41	4.21	-.19*

A low score indicates that a response is judged to be appropriate; high scores indicate that the response is judged inappropriate.

The effect size was calculated by dividing the mean difference (off-task minus on-task) by the standard deviation of the response variable.

*$p < .01$

**$p < .001$

not significantly different for on-task and off-task students. The judgments of prospective and experienced teachers were similar regarding how to respond to on-task vs. off-task students.

Gender, race, and ability. There were no main effects of students' gender or race on subjects' judgments about how to respond. Student ability, however, influenced subjects' judgments (Table 2). It was found that all main effects of ability were linear (i.e., no quadratic contrasts were statistically significant). Whenever the main effect of ability was significant, the linear contrast was also significant, simplifying interpretation of the ability effects.

Subjects were more likely to sustain low-ability than high-ability students, through providing a clue, $\underline{d} = .42$, $\underline{F}\ (1,204) = 76.44$, $\underline{p} < .001$, or by rephrasing the question, $d = .28$, $F\ (1,204) = 31.67$, $\underline{p} < .001$. Subjects were more likely to terminate high- than low-ability students, by asking another student to answer the question, $\underline{d} = -.16$, $\underline{F}\ (1,204) = 11.54$, $\underline{p} < .01$. They were more likely to provide negative feedback to highs rather than lows, $\underline{d} = -.19$, $\underline{F}\ (1,204) = 24.36$, $\underline{p} < .01$, more likely to provide positive feedback to lows than to highs, $\underline{d} = .21$, $\underline{F}\ (1,204) = 30.00$, $\underline{p} < .01$, and more likely to summarize for lows than for highs, $\underline{d} = .40$ $\underline{F}\ (1,204) = 69.60$, $\underline{p} < .001$. They were also more likely to ask another student to volunteer when they addressed lows than highs, $\underline{d} = .16$, $\underline{F}\ (1,204) = 10.89$, $\underline{p} < .01$, and more likely to provide no feedback to highs than to lows, $\underline{d} = -.14$, $\underline{F}\ (1,204) = 9.98$, $\underline{p} < .01$.

Teacher experience. Experienced teachers' judgments differed from those of prospective teachers in three ways: Experienced teachers were significantly

Table 2

Effects of Student Ability

	Linear effect size	Linear (dr = 1,204)	Nonlinear (dr = 1,204)
Sustaining behaviors			
Provide a clue	.42	76.44**	1.13
Rephrase the question	.28	31.67**	3.14
Summarize	.40	69.60**	.40
*Terminating behaviors			
Ask other student	-.16	11.54*	.70
Ask other to help	.16	10.89*	1.95
Feedback			
Positive feedback	.21	30.00*	1.86
Negative feedback	-.19	24.36*	1.47
No feedback	-.14	9.98*	.43

The linear contrast was computed by subtracting the "low ability" mean from the "high ability" mean.

* < .01

** < .001

more likely to choose to provide the answer themselves to the student responding incorrectly, $d = -.47$, $F\ (1,102) = 9.20$, $p < .01$; to ask a new question of the same student, $d = -.33$, $F\ (1,102) = 5.53$, $p < .05$; or to provide no feedback, $d = -.32$, $F\ (1,102) = 4.83$, $p < .05$.

Two-way Interaction Effects

Task-by-ability. A priori contrasts for the task-by-ability interaction effect were statistically significant for three response options: termination of the responding student by "asking another," $F\ (2,204) = 3.14$, $p < .05$; sustaining by providing a clue, $F\ (2,204) = 3.48$, $p < .05$; and sustaining by rephrasing the question, $F\ (2,204) = 10.06$, $p < .01$. In the latter two cases, the task-by-linear contrast was significant but not the task-by-quadratic contrast. Linear or quadratic contrasts failed to achieve significance on the other nine response options.

To clarify these latter two significant effects, recall that for these response options there had been a significant linear main effect of ability. The two-way interaction may be characterized as follows: this linear effect of ability was significantly more pronounced in the off-task than the on-task condition. Specifically, teachers' tendency to provide clues more for low- than for high-ability students was more pronounced in the off-task condition, $d = .54$, than in the on-task condition, $d = .30$ (Figure 1a). This difference in magnitudes of linear effect was statistically significant, $F\ (1,204) = 6.45$, $p < .05$.

Similarly, teachers' tendency to rephrase questions more for lows than for highs was particularly pronounced in the off-task condition, $d = .51$, but essentially nonexistent in the on-task condition, $d = .08$ (Figure 1b). This difference in magnitude of linear effect was also statistically significant,

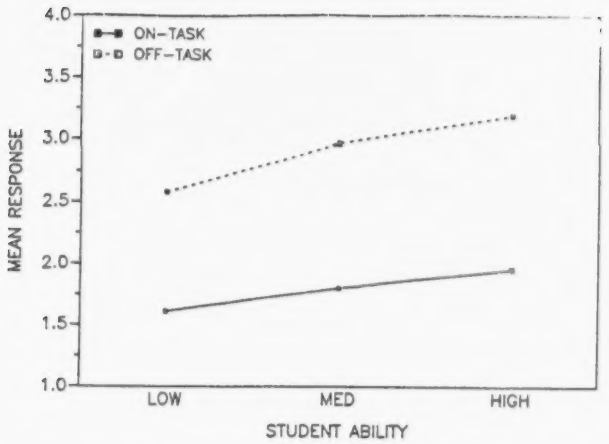

Figure 1a. Likelihood of providing a clue as a function of student ability and task behavior.

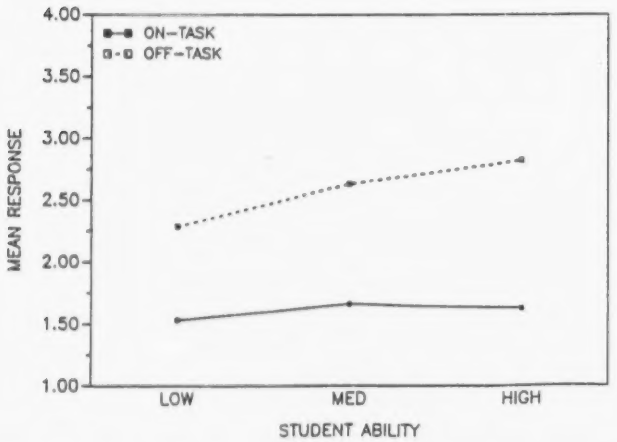

Figure 1b. Likelihood of rephrasing a question as a function of a student ability and task behavior.

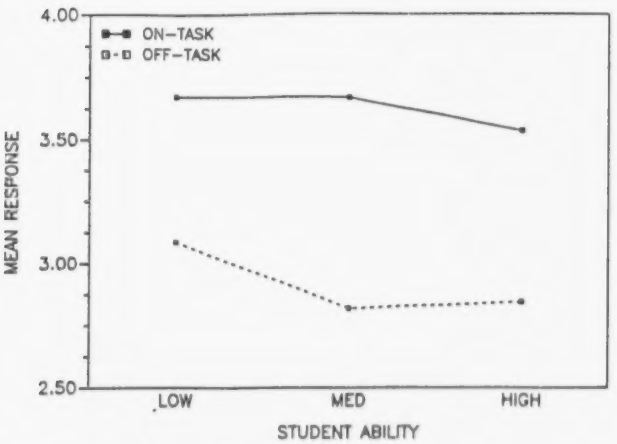

Figure 1c. Likelihood of asking another student as a function of student ability and task behavior.

F 1,204) = 20.14, $p < .001$. Another way to consider the same results is to say that, for these sustaining behaviors (providing clues and rephrasing), the effect of task behavior was more pronounced for high- than for low-ability students. Specifically, teachers were less sustaining toward off-task students overall, and this effect of task behavior was more pronounced for high- than for low-ability students.

The task-by-ability interaction was less readily interpretable in the case of the "ask other" option (Figure 1c). In this case, the task-by-linear contrast was not significant, but the task-by-quadratic contrast was significant, F (1,204) = 5.34, $p < .05$. The effect of task behavior appeared most pronounced in the middle ability condition. This result is related to the foregoing results in various ways. Although teachers were generally unlikely to recommend that on-task students be given terminating responses, they appeared more likely to recommend terminating interactions with on-task highs than with other on-task students.

In addition, although teachers were likely to recommend terminating responses for off-task students, they were less likely to make this recommendation for off-task lows than for other off-task students. Post hoc analysis using the Scheffé method indicated that the contrast between "highs" and others within the on-task condition approached but did not achieve statistical significance at the five percent level, $d = .12$, F (1, 204) = 4.46. However, the contrast between lows and others within the off-task condtion was significant, $d = .22$, F (1, 204) = 15.69, $p < .01$.

Other interactions with ability. For the race-by-ability effect, six a priori contrasts were statistically significant. The minority/majority-by-linear contrast was significant in the cases of the negative feedback and

clue options. Specifically, teachers' tendency to provide more negative feedback to high- than to low-ability students was more pronounced for ethnic majority children, $\underline{d} = -.29$, than for minority children, $\underline{d} = -.14$. The difference between these magnitudes of linear effect was statistically significant, $\underline{F}\ (1,204) = 5.72$, $\underline{p} < .05$. Teachers' tendency to provide more clues to low- than to high-ability students was more pronounced for minority youngsters, $\underline{d} = .55$, than for majority youngsters, $\underline{d} = .28$. Again, the difference between these magnitudes of linear effect was significant, $\underline{F}\ (1,204) = 5.61$, $\underline{p} < .05$.

Perhaps surprisingly, the black/Hispanic-by-quadratic ability contrast was statistically significant for four dependent measures: negative feedback, providing a clue, positive feedback, and summarizing. These effects are summarized below.

1. Negative feedback. Small quadratic effects of opposite sign yielded a significant black vs. Hispanic-by-quadratic interaction effect, $\underline{F}(1,204) = 13.25$, $\underline{p} < .01$. Because this result by itself was difficult to interpret, a post hoc analysis of simple main effects was performed. This analysis revealed that for Hispanic students, there was an effect of ability: highs were more likely targets of negative feedback than were other Hispanics, $\underline{d} = .21$, $\underline{F}(1,204) = 13.45$, $\underline{p} < .01$. However, there were no significant ability effects for blacks.

2. Provision of clues. Again, small quadratic effects of opposite sign yielded a significant black vs. Hispanic-by-quadratic interaction effect. Post hoc results showed that for Hispanics, highs were significantly less likely than others to receive clues, $\underline{d} = .49$, $\underline{F}(1,204) = 30.32$, $\underline{p} < .01$. The Hispanic average and low conditions were similar. For blacks, lows were

significantly more likely than others to receive clues, $\underline{d}$ = .34, $\underline{F}(1,204)$ = 23.34, $\underline{p} < .01$. The black high and average conditions were similar.

3. Positive feedback. The black vs. Hispanic-by-quadratic interaction resulted from a pattern similar to that described for clues. For Hispanics, highs were less likely to receive positive feedback than others, $\underline{d}$ = .19, $\underline{F}(1,204)$ = 11.40, $\underline{p} < .05$, and average and high Hispanic conditions were similar. For blacks, lows were more likely to receive positive feedback than others, $\underline{d}$ = .23, $\underline{F}(1,204)$ = 16.53, $\underline{p} < .01$, and average and high conditions were similar.

4. Summarizing. The black vs. Hispanic-by-quadratic interaction followed a similar pattern to that of clues and positive feedback. For Hispanics, summarizing was less likely for highs than for others, $\underline{d}$ = .46, $\underline{F}(1,204)$ = 41.96, $\underline{p} < .01$, and middle and low conditions were similar. For blacks, summarizing was more likely for lows than for others, $\underline{d}$ = .32, $\underline{F}(1,204)$ = 19.75, $\underline{p} < .01$; average and high conditions had similar means.

Significant gender-by-ability interaction effects were found for negative feedback, rephrase, clue, positive feedback, summarize, and volunteer. In each case, a significant gender-by-quadratic effect was found, and in each case the interaction effects were significant because small quadratic effects had opposite signs for boys and girls. One significant gender-by-linear ability effect was found: There was a linear effect of ability for girls, $\underline{d}$ = .34, but not for boys, $\underline{d}$ = .00. Specifically, teachers were more likely to ask another student to help low-ability girls than high-ability girls. This effect of ability was absent for boys in both high- and low-ability categories.

Other two-way interactions. There were four additional significant two-way contrasts. First, the race-by-gender interaction was marginally significant for negative feedback, F (2,102) = 4.04, $p < .05$. However, the post hoc analysis showed that gender effects were small and nonsignificant within each condition of race. The statistical significance of the interaction resulted from the different directions of these nonsignificant effects. For whites, the difference between male and female means was slightly negative, d = -.12; for Hispanics, it was slightly positive, d = .09; for blacks it was near zero, d = -.03.

Second, there was a race-by-task effect on the asking a new question response. The effect of task behavior was more pronounced for blacks, d = -.34, than for Hispanics, d = .05. Specifically, for blacks, being on task increased the likelihood of a new question, but for Hispanics, that likelihood was similar for on-task and for off-task students.

Third, gender-by-task interactions were found for negative feedback, positive feedback, and no feedback. In each case, the effect of task behavior was somewhat more pronounced for boys than for girls. Specifically, teachers' tendency to provide more negative feedback to off-task than to on-task students was more pronounced for boys, d = 1.17, than for girls, d = 1.06, F (1,102) = 4.62, $p < .05$.; teachers' tendency to provide more positive feedback for on-task than for off-task students was more pronounced for boys, d = -1.05, than for girls, d = -.91, F(1,102) = 7.68, $p < .05$. The increased likelihood of no feedback for off-task as opposed to on-task behavior was more pronounced for boys, d = .27, than for girls, d = .13, F(1,102) = 4.39, $p < .05$.

Finally, in the case of asking another to volunteer, the effect of task behavior was greater for experienced than for inexperienced teachers.

Experienced teachers judged this response more appropriate for off-task than for on-task students, $\underline{d} = .23$, while a smaller effect of task appeared for inexperienced teachers, $\underline{d} = .11$. The difference between these effect sizes was significant, $\underline{F}(1,102) = 7.19$, $\underline{p} < .01$.

Higher Order Interactions

For each of the 12 response options, the 16 possible higher order interactions were tested. Of these 192 interaction effects, 28 were statistically significant at the 5% level, certainly more than could have been expected on the basis of chance. However, it seemed essential to use extreme caution in interpreting these higher order interaction effects, both because of the possibility of Type I errors and also because the available theory for interpretation cannot accommodate the complexity of these interactions. Below are those higher order interaction effects which both (a) follow a clear pattern and (b) hold some promise of aiding interpretation of results discussed above. Two types of interactions fulfilled these conditions: gender-by-race-by-task and ability-by-task-by-race.

1. Gender-by-race-by-task. This interaction was significant for negative feedback, positive feedback, providing clues, and rephrasing. In each case the pattern was the same: There was a gender-by-task effect for whites, but not for blacks or Hispanics. In each case, the effect of task behavior was more pronounced for white boys than for white girls.

2. Ability-by-task-by-race. This interaction was statistically significant in the case of negative feedback, positive feedback, rephrasing, and summarizing. It was noted earlier that the race-by-quadratic interaction was significant for each of these response variables. Post hoc analysis of

the three-way interaction revealed that this race-by-quadratic interaction, discussed in detail earlier, appeared only in the off-task condition for each of these four outcomes.

Discussion

Two recurrent findings in classroom research provide the empirical context for the present study. First, evidence has accumulated which indicates that teacher behaviors which maximize student engagement also tend to reflect positive expectations for student performance and to increase student achievement. Second, numerous studies suggest that teachers often tend to differentiate their use of these behaviors in ways which may favor some students and be detrimental to others, especially minority youngsters, girls, and those perceived to have low ability. Since classroom life is complex and fast paced, teachers experience many conflicting demands. Thus, uncovering and explaining the sources of discrete teacher actions is a formidable challenge. The present study focused on the antecedents of one important source for teacher action: teachers' judgments about the appropriateness of responses to individual students during public recitation. Although the conflicting demands of classroom life may ultimately require many compromises of pedagogical judgment, we presume that the initial bases of teacher judgment must be understood before teacher action becomes comprehensible.

Summary and Interpretation of Key Findings

Main effects. The most important determinant of teachers' judgments about appropriate responses in this study was the behavior of the student responding incorrectly to the question. Teachers were substantially more likely to recommend sustaining through providing clues, rephrasing the question, or

summarizing for on-task students than for off-task students. Teachers were substantially more likely to opt for terminating off-task students by asking another student to answer the question. They were also more likely to opt for providing off-task students more negative feedback and less positive feedback and were more likely to recommend no feedback for off-task students. These findings show that both practicing and prospective teachers judge as highly appropriate the differentiation of their feedback to students on the basis of those students' efforts to engage the content of the lesson. Such differentiation is consistent with current research on classroom interaction.

The main effect of student ability was also consistent. Though the magnitudes of the ability effects were smaller than the behavior effects, they were significant nevertheless. Teachers were inclined to opt for sustaining low-ability students more than high-ability students through the use of clues, rephrasing, positive feedback, or seeking a volunteer to help. In contrast, they tended to choose to terminate responses of high-ability students who gave incorrect answers by asking another student to answer the question or by providing negative feedback or no feedback. Thus, the results indicate virtually the same patterns of sustaining treatment for on-task students generally as for low-ability students specifically.

The main effect of ability in this study provides an interesting and perhaps surprising contrast to prior research on teacher expectations. Research suggests that teachers tend to sustain high-ability students more than low-ability students during public recitations. The results of the present study suggest that the teachers viewed sustaining as more appropriate for lows than for highs. Several explanations for this apparent contradiction are possible.

First, it could be that the subjects in the present study are fundamentally different in their judgments from teachers in past studies. This speculation does not seem warranted, however, because of the finding that prospective and experienced teachers were similar in their responses to students of differing ability. Second, it could be that the judgments of subjects in this hypothetical study do not reflect subjects' judgments during actual classroom situations. This possibility will be investigated based on classroom observation data currently being collected.

A third possibility is that teachers sustain students for different reasons and that sustaining may convey different messages depending on the context and purpose of its use. Teacher expectancy literature suggests, for example, that teachers sustain students because they expect the students to produce the correct answer, thus facilitating progress in mastering the content of the lesson (Cooper, 1977). Similarly, terminating responses may reflect a belief that to elicit the correct answer the teacher must move to a "better" student who can be counted on to provide it. A different motivation for sustaining is to provide extra help to those who stand most to benefit from it--that is, to "concern students" in the language of Silberman (1969, cited in Brophy and Evertson, 1981). The function for this kind of sustaining might be, as Brophy (1983) suggests, to compensate for the problematic behavior of such students and to maximize their chances of achieving success. This idea helps provide a unifying explanation for the effects of task behavior and ability in the study: On-task students merit sustaining and low-ability students need it.

Interactions with task behavior. Subjects in this study strongly and appropriately differentiated their responses on the basis of student's task

behavior. The task-by-ability interaction revealed, however, that this differentiation was significantly more pronounced for high-ability than for low-ability students. That is, the probability of a sustaining response was highly contingent on the task behavior of highs and significantly less so for lows. A disconcerting implication of this result is that the differentiated feedback accorded highs is more likely to encourage "attribution to effort" (Dweck & Goetz, 1978) than the less differentiated feedback for lows. Despite the view that teachers may see lows as "concern students," sustained interaction appears to have less value for them since they did not have to be engaged to be sustained. Thus, the message may be one of attribution to ability, suggesting that low-ability students are provided sustaining responses by virtue of being low rather than as a result of their efforts.

The interactions between student gender and task behavior for positive, negative, and no feedback are consistent with the findings from previous research (e.g., Bossert, 1981). In general, boys received more differentiated feedback of all three types. The fact that this gender-by-task interaction was restricted to whites suggests, perhaps, the special salience of white boys in the classroom.

Interactions with race. Results for the race-by-ability effect indicate that the teachers' tendency to provide more negative feedback to highs than to lows was more pronounced for majority than minority students. Subsequent post hoc analyses of the black/Hispanic-by-quadratic ability contrast further revealed that there was a significant effect of ability for white and Hispanic students but not for black youngsters. While differential treatment by race is clearly indicated, the particular meaning of negative feedback is less clear. To some teachers in the study, "negative" may have meant appropriate corrective

feedback on the student's wrong answer rather than criticism of the student's answer or effort or more "negative" treatment based on student race. Future studies incorporating follow-up interviews with simulations would help to explain the implicit bias found in this study.

An unanticipated but interesting pattern appeared in the race-by-task-by-ability interaction. Among Hispanics, highs were significantly less likely than middle- or low-ability students to receive positive feedback, clues, or summaries. Among blacks, low-ability students were more likely than middle- or high-ability students to receive each of these "sustaining" responses. The overall significance of these interaction effects derives entirely from their presence in the off-task condition. If the sustaining of low-ability students is an expression of concern (in Silberman's terms), then subjects in this study had great "concern" for low-ability blacks, less pronounced "concern" for middle-ability blacks, high-ability blacks, low-ability Hispanics and middle-ability Hispanics, and little "concern" for high-ability Hispanics.

Effect of teacher experience. The results for teacher experience indicate that experienced teachers are more likely to terminate interaction by answering questions themselves, asking a new question, or providing no feedback when a student answers incorrectly. These findings suggest a desire to move the lesson along and are perhaps more structural than psychological. In essence, these judgments of experienced teachers probably closely resemble their actions in their own classrooms where pedagogical judgments and expectancies may be circumscribed by the press of moment-to-moment classroom life. What is perhaps most striking is the unexpected absence of significant differences between experienced and prospective teachers. While the prospective teacher's "apprenticeship of observation" (Lortie, 1975) as a student in K-12 classrooms

might, in some ways, serve as a surrogate for a practicing teacher's experience, the results of this study indicate that both prospective and experienced teachers nonetheless made similar judgments which led to differential student treatment under hypothetical circumstances. The similarities between the two groups more likely suggest a shared social ideology rather than a view clearly explainable by teaching experience in real classrooms.

Conclusions

Although the policy-capturing method used in this study is not designed to uncover specific contextual features of the classroom or to reflect the cognitive processes subjects used to make their decisions, these results do suggest five conclusions about teachers' pedagogical judgments during classroom interaction. First, teachers strongly differentiated their recommended responses on the basis of students' task behavior. Second, student ability significantly influenced teacher judgments. Low-ability students were treated "like" on-task students in that they were more likely targets of positive, sustaining behavior than high-ability students. Third, the evidence suggested that although low-ability students were "like" on-task students in receiving sustaining responses, the significance and meaning of sustaining appeared to be different in the two cases. Since sustaining was less contingent on behavior for low-ability students than for high-ability students, such sustaining response seems less an encouragement of effort than an expression of concern for success. More research is needed to discern between both teacher intentions for sustaining behavior and the manner in which students interpret such messages. Fourth, there were no main effects of student gender or race/ethnicity on teacher judgments. Nevertheless,

interactions of these factors with task behavior and abilty suggest that subjects interpreted information about ability and behavior differently for boys than for girls and differently for different racial/ethnic groups. The content, causes, and consequences of such differentiation supply motivation for further research. Finally, the judgment profiles of prospective and experienced teachers were remarkably similar, reflecting perhaps shared cultural or ideological views rather than effects of age, training, or classroom experience.

One must question the appropriateness of the nonbehavioral bases for differentiation of teacher behavior found in this study. Such bases for differential teacher treatment are likely to lead to behaviors that convey messages which can have deleterious effects on student performance and outcomes. In contrast, differential treatment of students may be regarded as appropriate and necessary only when teachers conceptualize ability, as Marshall and Weinstein (1984) suggest, as a "multidimensional" configuration to be judged on the basis of performance and developed on the basis of need, when they convey the value of individual differences and when they strive to create opportunities for all students to learn. Incorrect student responses are essentially public displays of student incompetence. How teachers respond determines the messages communicated to students about their potential for academic achievement and intellectual growth.

References

Aaron, R., & Powell, G. (1982). Feedback practices as a function of teacher and pupil race during reading groups instruction. Journal of Negro Education, 51, 50-59.

Adams, G. R., & LaVoie, J. (1974). The effect of students' sex, conduct, and facial attractiveness on teacher expectancy. Education, 95, 76-83.

Aloia, G., Maxwell, J., & Aloia, S. (1981). Influence of a child's race and the EMR label on initial impressions of regular classroom teachers. American Journal of Mental Deficiency, 85, 619-623.

Argulewicz, E., Elliott, S., & Hall, R. (1982). Comparison of behavioral ratings of Anglo-American and Mexican-American gifted children. Psychology in the Schools, 19, 469-472.

Baron, R. M., Tom, D., & Cooper, H. M. (1985). Social class, race and teacher expectancies. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), Teacher expectancies (pp. 251-269). Hillsdale, NJ: Erlbaum.

Blumenfeld, P., Hamilton, V., Wessels, K., & Falkner, D. (1977). 'You can,' 'you should,' and 'you'd better': Teacher attributions regarding achievement and social behavior. Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Bock, R. D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw-Hill.

Bossert, S. (1981). Understanding sex differences in children's classroom experiences. Elementary School Journal, 81, 255-266.

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. Journal of Educational Psychology, 75, 633-661.

Brophy, J., & Evertson, C. M. (1981). Student characteristics and teaching. New York: Longman.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and school achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Brophy, J., & Good, T. (1974). Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart & Winston.

Cohen, J. (1969). Statistical power analyses for the behavioral sciences. New York: Academic Press.

Cooper, H. M. (1977). Controlling personal rewards: Professional teachers' differential use of feedback and the effects of feedback on the student's motivation to perform. Journal of Educational Psychology, 69, 419-427.

Cooper, H. M., & Good, T. L. (1983). Pygmalion grows up: Studies in the expectation communication process. New York: Longman.

Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. Curriculum Inquiry, 14, 129-150.

Dusek, J. B., & Joseph, G. (1985). The bases of teacher expectancies. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), Teacher expectancies (pp. 229-250). Hillsdale, NJ: Erlbaum.

Dweck, C. S., & Goetz, T. E. (1978). Attributions and learned helplessness. In R. Kidd (Ed.), New directions in attribution research, (Vol. 2, pp. 158-179). Hillsdale, NJ: Erlbaum.

Etaugh, C., & Harlow, H. (1975). Behaviors of male and female teachers as related to behaviors and attitudes of elementary school children. Journal of Genetic Psychology, 127, 163-170.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn (pp. 7-32). Washington, DC: U. S. Government Printing Office.

Floden, R. E., Porter, A. C., Schmidt, W., Freeman, D., & Schwille, J. (1981). Responses to curriculum pressures: A policy-capturing study of teacher decisions about content. Journal of Educational Psychology, 73, 129-141.

Good, T. (1970). Which pupils do teachers call on? Elementary School Journal, 70, 190-198.

Good, T., & Findley, M. (1985). Sex role expectations and achievement. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), Teacher expectancies (pp. 271-300). Hillsdale, NJ: Erlbaum.

Irvine, J. J. (1985). Teacher communication patterns as related to the race and sex of the student. Journal of Educational Research, 78, 338-345.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks-Cole.

Leinhardt, G., Seewald, A. M., & Engel, M. (1979). Learning what's taught: Sex differences in instruction. Journal of Educational Psychology, 71, 432-439.

Lortie, D. C. (1975). Schoolteacher: A sociological study. Chicago: University of Chicago Press.

Marshall, H. H., & Weinstein, R. S. (1984). Classroom factors affecting students' self-evaluations: An interactional model. Review of Educational Research, 54, 301-325.

Mendoza, S., Good, T., & Brophy, J. (1972). Who talks in junior high classrooms? (Report No. 68). Austin: University of Texas, Research and Development Center for Teacher Education.

Purgess, P. (1979). Teacher expectancy for academic success in relation to label, sex, and pupil behavior. Dissertation Abstracts International, 39, 2003-2004A.

Roberts, T., Hutton, J., & Plata, M. (1985). Teacher ratings of Hispanic, black and Anglo students' classroom behavior. Psychology in the Schools, 22, 353-356.

Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), The psychology of teaching methods (75th yearbook of the National Society for the Study of Education, Part I, pp. 335-371). Chicago: University of Chicago.

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. Behavioral and Brain Sciences, 3, 377-386.

Rowe, M. (1974). Wait-time and rewards as instructional variables, their influence on language, logic and fate control: Part One--Wait-time. Journal of Research in Science Teaching, 11, 81-94.

Seber, G. A. F. (1977). Linear regression analysis. New York: Wiley.

Silberman, M. (1969). Behavioral expression of teachers' attitudes toward elementary school children. Journal of Educational Psychology, 60, 402-407.

Simpson, A. W., & Erickson, M. T. (1983). Teachers' verbal and nonverbal communication patterns as a function of teacher race, student gender, and student race. American Educational Research Journal, 20, 183-198.

Smith, M. L. (1980). Teacher expectations. Evaluation in Education, 4, 53-55.

Willis, S., & Brophy, J. (1974). The origins of teachers' attitudes toward young children. Journal of Educational Psychology, 66, 520-529.

Zahorik, J. A. (1968). Classroom feedback behavior of teachers. Journal of Educational Research, 62, 147-150.